



## Evaluating Mineral Lattices as Evolutionary Proxies for Metalloprotein Evolution

Kenneth N. McGuinness<sup>1</sup> · Gunnar W. Klau<sup>2</sup> · Shaunna M. Morrison<sup>3</sup> · Elisha K. Moore<sup>4</sup> · Jan Seipp<sup>2</sup> · Paul G. Falkowski<sup>5</sup> · Vikas Nanda<sup>1</sup>

Received: 22 December 2021 / Accepted: 3 October 2022 / Published online: 16 November 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

### Abstract

Protein coordinated iron-sulfur clusters drive electron flow within metabolic pathways for organisms throughout the tree of life. It is not known how iron-sulfur clusters were first incorporated into proteins. Structural analogies to iron-sulfide minerals present on early Earth, suggest a connection in the evolution of both proteins and minerals. The availability of large protein and mineral crystallographic structure data sets, provides an opportunity to explore co-evolution of proteins and minerals on a large-scale using informatics approaches. However, quantitative comparisons are confounded by the infinite, repeating nature of the mineral lattice, in contrast to metal clusters in proteins, which are finite in size. We address this problem using the Niggli reduction to transform a mineral lattice to a finite, unique structure that when translated reproduces the crystal lattice. Protein and reduced mineral structures were represented as quotient graphs with the edges and nodes corresponding to bonds and atoms, respectively. We developed a graph theory-based method to calculate the maximum common connected edge subgraph (MCCES) between mineral and protein quotient graphs. MCCES can accommodate differences in structural volumes and easily allows additional chemical criteria to be considered when calculating similarity. To account for graph size differences, we use the Tversky similarity index. Using consistent criteria, we found little similarity between putative ancient iron-sulfur protein clusters and iron-sulfur mineral lattices, suggesting these metal sites are not as evolutionarily connected as once thought. We discuss possible evolutionary implications of these findings in addition to suggesting an alternative proxy, mineral surfaces, for better understanding the coevolution of the geosphere and biosphere.

**Keywords** Iron-sulfur clusters · Protein evolution · Greigite · FeS minerals · Mineral evolution · Ferredoxin · Protein-mineral co-evolution

✉ Kenneth N. McGuinness  
kenneth.mcguinness@gmail.com

<sup>1</sup> Center for Advanced Biotechnology and Medicine, Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA

<sup>2</sup> Algorithmic Bioinformatics Group, Department of Computer Science, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>3</sup> Earth and Planets Laboratory, Carnegie Institute for Science, Washington, D.C., USA

<sup>4</sup> Department of Environmental Science, Rowan University, Glassboro, NJ, USA

<sup>5</sup> Environmental Biophysics and Molecular Ecology Program, Department of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA

## Introduction

Proteins coordinating multinuclear iron-sulfur (FeS) clusters are abundant in all domains of life (Andreini et al. 2007). They potentiate the multitude of electron transfer reactions essential for metabolism (Williams 1981). FeS proteins are likely among the most ancient, and appear at the root of primary metabolic pathways presumed to be present on early Earth (Raanan et al. 2018, 2020). The origin of protein binding FeS cluster formation on early Earth is unknown. The cubane 4Fe4S clusters found in central proteins such as bacterial ferredoxin, are qualitatively similar to FeS minerals such as greigite ( $\text{Fe}(\text{II})\text{Fe}(\text{III})_2\text{S}_4$ ), an ancient mineral found near hydrothermal vents (Russell and Hall 1997; Rickard and Luther 2007; Zhao et al. 2020). This has led to the idea that FeS sites in proteins originated from an ancestral peptide-mineral interaction, where the mineral surface was capable of prebiotic chemical reactions similar to those in biotic metabolisms (de Aldecoa et al. 2013; Nitschke et al. 2013; Novikov and Copley 2013; Roldan et al. 2015). While comparisons of proteins and minerals have often been anecdotal and qualitative, extensive databases of proteins (i.e., the PDB (Berman 2000)) and minerals (i.e., the AMCSd (Downs and Hall-Wallace 2003)) make it possible to use chemoinformatic approaches to quantitatively assess chemical similarity between metal sites in minerals and proteins, and test models of geosphere and biosphere co-evolution. However, few such studies have been attempted, due in part to the challenge of comparing a repeating mineral lattice to the finite protein metal site.

Recently, Zhao et al. (2020) approached this problem of quantifying protein-mineral metal cluster similarity and pioneered the use of a cheminformatics molecular fingerprint approach. To avoid issues with mineral unit cell and protein cluster size comparisons (i.e., infinite vs. finite), the protein cluster was aligned by iterative rotation and translation to an extended three-dimensional mineral lattice. Tantalizing topological similarities were observed between FeS minerals like greigite and the protein 4Fe-4S clusters such as those found in ferredoxin. However, the Fe coordination is distinct, octahedral in greigite and tetrahedral in ferredoxins. An approach that considers both topological and chemical measurements would be useful in quantifying protein-mineral similarity. We develop a graph-based method to directly find similar substructures that share the same structural topology (i.e., coordination number and cluster dimensionality).

In this paper, we address the problem of comparing an infinite mineral lattice to a finite protein metal cluster in two steps. We used the Niggli reduction (Křivý and Gruber 1976) to simplify the mineral unit cell into a unique finite structure that, when translated, reproduces the entire crystal lattice. Next, we developed and applied a graph-theoretic approach to finding the maximum common connected edge subgraph (MCCES) of the reduced mineral unit cell quotient graph and protein metal cluster graph. Graph similarity was computed using the Tversky index (Tversky 1977) between subgraphs to account for differences in graph size. Directly comparing clusters as graphs avoided costly iterative structure comparison calculations, and did not require transforming the protein cluster to fit the mineral lattice, as was previously performed (Zhao et al. 2020). Representing the atoms as nodes and bonds as edges easily allowed us to include additional chemical attributes such as atom coordination number. In addition to comparing cluster similarity, we also compared mineral and protein cluster iron coordination properties, including first shell metal–ligand bond lengths, bond angles, and deviation from ideal coordination geometry. We find that FeS mineral lattice environments differ from that of protein metal site environments. Mineral lattices are relatively rigid, whereas protein metal sites are more flexible. We suggest that

Fe-S mineral lattices are not ideal proxies for studying the co-evolution of the biosphere and geosphere. Alternatively, we suggest the topologically diverse and chemically active mineral surface may prove to be a better proxy for understanding the evolutionary relationships between minerals and proteins.

## Materials and Methods

**Datasets** Protein iron-sulfur sites were downloaded from the MetalPDB (Putignano et al. 2018) and filtered based on three dimensional structure resolution of  $\leq 1.7 \text{ \AA}$  and sequence identity less than 90%. Structures were filtered based on a previous study showing anomalies increase with poorer structural resolution, and that structures with resolution between 1.5 and 1.8  $\text{\AA}$  may be the minimal resolution required for proper metal-assignment and reliable analysis of metal site geometric environments (Müller et al. 2003). A list of named protein structures and sites used is provided in the Supplementary Information. There were 32 iron-sulfur mineral structures in the American Mineral Crystallography Structural Database (AMCSD) accessed in November 2021. These included duplicates of several minerals with structures solved under diverse laboratory conditions. One representative structure was chosen for each mineral species, based on the criteria of single atom occupancy (i.e., only containing Fe and S atoms), and laboratory conditions closest to standard temperature and pressure. These include AMCSD ID# 127 (greigite), 12,728 (pyrite), 288 (pyrrhotite), 14,518 (mackinawite), 12,726 (marcasite), 80 (smythite), 15,564 (töchilinite), and 4160 (troilite). The AMCSD database was chosen instead of other inorganic material databases (e.g., the Inorganic Crystal Structure Database (Zagorac et al. 2019)) because of its geographic focus.

**Iron Coordination Statistics** For each metal coordination environment, the geometry and geometric root mean square deviation (gRMSD), from ideal geometry, was calculated using the tool FindGeo (Andreini et al. 2012). FindGeo superimposes the metal-coordinating atoms of each coordination environment to a library of templates with alternative ideal geometries, and the best geometry is chosen by the lowest RMSD from the ideal structure. Bond length and angle were computed using a Biopython (Cock et al. 2009) script.

**Niggli-Reduced Unit Cell Generation** As mentioned previously, the arrangement of atoms in the mineral crystal lattice defines an infinite network of atoms, making direct mathematical comparisons between minerals and the finite molecular structure of a protein metal cluster a unique challenge. To properly make a comparison, a mineral lattice can be represented as a unit cell with finite size and geometry. However, there may be an infinite number of unit cells (i.e., primitive cells) that can equally represent a three-dimensional periodic lattice. The Niggli-reduced unit cell, one of infinitely many unit cells, uniquely describes the crystal lattice independent of lattice symmetry. (Santoro and Mighell 1970) For this study, the STRUCTURE TIDY (Gelato and Parthé 1987) Niggli-Reduced Cell utility, implemented in Vesta (version 3.4.7 (Momma and Izumi 2011)) was used to transform the original unit-cell parameters, of each AMCSD mineral structure, to the parameters of the Niggli-reduced unit cell. For this study, the Niggli reduction method was chosen over other reduction methods (e.g., Selling reduction (Andrews et al. 2019)) because it guarantees convergence to a unique reduced cell for all Bravais lattices. (Santoro and Mighell 1970).

**Graph Generation** A mineral crystal structure is an infinite and periodic network (i.e., a crystallographic net (Klee 2004)) that can be viewed as a graph consisting of atoms as nodes and chemical bonds as edges. Due to crystallographic net periodicity, each net can be transformed into a finite graph where all translationally equivalent nodes of the net are projected onto one node of a finite graph. In other words, the nodes of the resultant finite graph represent only translationally non-equivalent nodes and edges of each net. The resultant finite graph is called the quotient graph of the crystallographic net (Chung et al. 1984; Klee 2004). A labeled quotient graph completely determines the topology of a crystallographic net. Previous studies successfully projected Niggli-reduced until cells of mineral crystal lattices onto their respective quotient graphs to quantitatively measure mineral topological complexity (Krivovichev 2012a, b, 2013; Krivovichev et al. 2018). Following previous studies (Krivovichev 2013; Krivovichev et al. 2018), Niggli-reduced unit cells of mineral crystal structures were projected onto quotient graphs using ToposPro version 5.4.1 (Blatov et al. 2000). Graphs of protein metal clusters were generated from the atom and chemical bond information in the structure.

**Metal Cluster Similarity and Maximum Common Connected Edge Subgraph Calculation** We developed an algorithm to compute the maximum common connected edge subgraph (MCCES) between the non-periodic protein cluster graph and the periodic mineral quotient graph. Our method, called como (code available at <https://github.com/AIBi-HHU/como>), is based on an integer linear programming formulation for the maximum common edge subgraph problem (Bahiense et al. 2012), to which we added connectivity constraints defined on the line graphs of both molecular graphs. We also added optional functionality to restrict the atom mappings to labels (e.g., identical coordination values). We used the number of MCCES edges to compute the Tversky index (Tversky 1977), which generalizes the Sorensen-Dice index and is more appropriate for local matching. The Tversky index was computed as,

$$\text{Tversky index} = \frac{\# \text{ shared edges}}{(\# \text{ shared edges} + \# \text{ edges not shared})}$$

where number of edges not shared are those within the protein cofactor. We implemented the algorithms in Python, using the networkx (Hagberg et al. 2008) library and used Gurobi 9.1 (Optimization 2021) to solve the underlying mathematical optimization problem. Mineral quotient graphs and protein metal cluster graphs were converted into node and edge files that were used as input files to como. Retaining chemical in addition to topological information within each graph, each node label contained the element name, atomic coordination number, and crystallographic dimensionality (i.e., 2D- or 3D-periodic net) were used as connectivity constraints.

## Results

**Fe-S Bond Lengths** Electrons within the first shell of a Fe atom (i.e.,  $< 2.8 \text{ \AA}$ ) directly affect the chemical properties of the Fe atom. For instance, increasing the bond strength between the Fe and cysteine sulfur directly impacts protein function.(Mathé et al. 2007) Therefore, comparing the first-shell properties of the Fe coordination center of protein and mineral clusters is an important step toward understanding if they are similar.

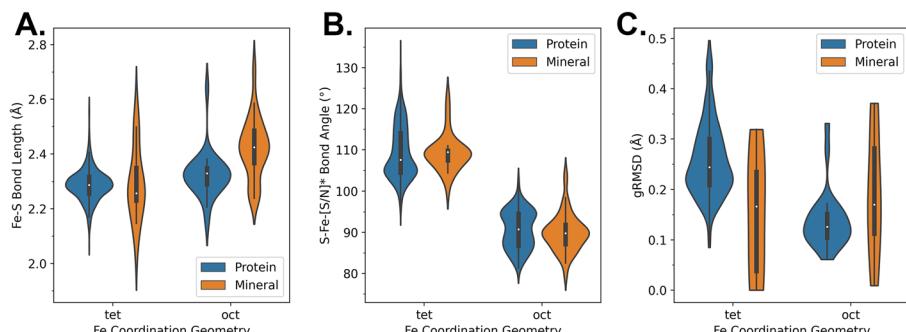
**Table 1** Comparison of mineral, protein tetrahedral and octahedral coordinated average ( $\pm$  standard deviation) iron bond angles and lengths and gRMSD from ideal geometry

Fe Coordination #	Environment	Fe-S (Å)	S-Fe-[S/N*] (°)	gRMSD (Å)
4	protein	2.287 (0.057)	109.30 (6.44)	0.26 (0.07)
	mineral	2.292 (0.132)	109.34 (4.45)	0.15 (0.13)
6	protein	2.323 (0.084)*	90.64 (4.50)*	0.14 (0.06)*
	mineral	2.422 (0.109)	89.88 (4.60)	0.19 (0.12)

\*protein octahedral Fe coordination is measured in the  $\text{FeN}_5\text{S}$  environment and for the angle S-Fe-N

Average tetrahedrally coordinated Fe-S bond lengths for minerals and proteins are essentially the same (Table 1) and are in agreement with previous studies (Zheng et al. 2017; Moriarty and Adams 2019). However, minerals have a larger range (i.e., min – max) of Fe-S bond lengths than proteins (2.042 – 2.582 Å vs. 2.061 – 2.578 Å; Fig. 1A). Given that ligand coordination number is an important factor governing metal-ligand distances (Kuppuraj et al. 2009), additional differences are possibly due to sulfur generally coordinating 2 or 3 atoms in biology, as compared to sulfur in minerals coordinating four or five other atoms. All-sulfur ligated octahedral iron coordination sites do not exist within proteins (Moriarty and Adams 2019) and therefore  $\text{FeN}_5\text{S}$  octahedral coordination sites of proteins (i.e., Fe coordination to heme and histidine nitrogen, and to the sulfur of methionine) were compared to all octahedral  $\text{FeS}_6$  mineral environments. Octahedral Fe-S coordination sites in minerals have a higher average bond length, compared to octahedral Fe-S sites within protein. However, both protein and mineral octahedral Fe-S are within a standard deviation,  $2.422 \pm 0.109$  Å vs  $2.321 \pm 0.08$ , respectively. Although a small number of protein octahedral Fe-S bonds exist in the dataset (24), the average Fe-S bond length are larger than and within a standard deviation of Fe-S bond lengths found in tetrahedral environments (Table 1; Fig. 1A).

**S-Fe-S Bond Angles** Mean S-Fe-S angles of tetrahedrally coordinated Fe are similar between proteins and minerals with proteins having a wider-angle range (i.e., min/max),  $94.80 - 133.56^\circ$  and  $100.04 - 123.34^\circ$ , respectively. Mean S-Fe-S angles for both proteins



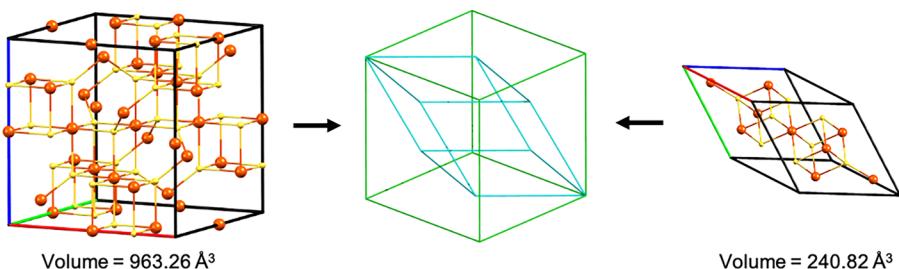
**Fig. 1** Comparison of tetrahedral (tet) and octahedral (oct) coordination of iron in protein and mineral **A** Fe-S bond lengths, **B** S-Fe-S bond angles, and **C** Geometric root mean standard deviation from ideal geometry. \*Note octahedral Fe coordination in proteins is within the  $\text{FeN}_5\text{S}$  coordination environment, panel **B** is represented as angle S-Fe-N for octahedral Fe coordination within proteins

and minerals are within the range of the ideal tetrahedral coordination angle ( $109.47^\circ$ ; Fig. 1B). Since proteins do not have octahedral Fe coordination sites containing all-sulfur ligands, as do minerals, octahedral Fe coordination sites within proteins containing N and S were compared to all sulfur ligated octahedral Fe coordination sites. Mean S-Fe-S and N-Fe-S angles were similar for minerals ( $89.88 \pm 4.60^\circ$ ) and proteins ( $90.65 \pm 4.50^\circ$ ) and are within range of ideal octahedral coordination ( $90^\circ$ ). Minerals have a similar minimum and maximum octahedral range of angle values than proteins ( $79.40 - 104.61^\circ$  vs.  $81.26 - 101.88^\circ$ ). (Fig. 1B).

**gRMSD** Protein tetrahedral Fe coordination deviates from ideal geometry to a greater extent than minerals, with gRMSD values of  $0.26 \pm 0.09$  and  $0.15 \pm 0.13$ , respectively. The minerals greigite ( $0.00 \text{ \AA}$ ), mackinawite ( $0.01 \text{ \AA}$ ), and tochnilinite ( $0.12 \text{ \AA}$ ) have gRMSD values closest to ideal tetrahedral geometry, whereas all protein tetrahedral coordination gRMSD values are above  $0.085 \text{ \AA}$  from an ideal geometry. Tochnilinite, has three Fe tetrahedral sites above  $0.2 \text{ \AA}$  ( $0.22 \text{ \AA}$ ,  $0.28 \text{ \AA}$ ,  $0.32 \text{ \AA}$ ), which is possibly due to crystallography artifacts arising from resolution issues (Organova et al. 1973). Average deviation from ideal octahedral Fe coordination is higher for minerals ( $0.19 \text{ \AA}$ ) than proteins ( $0.14 \text{ \AA}$ ). (Fig. 1C).

**Fe-S Cluster Analysis** Protein function is also affected by secondary shell interactions (Andreini et al. 2009; Valasatava et al. 2018) and therefore comparing only first-shell metal coordination interactions is insufficient. Unlike previous results (Zhao et al. 2020), we compare the entire Fe-S cluster environment within proteins, including the cysteine sulfur, when calculating mineral-protein cluster similarity. We focused on 4Fe4S clusters as they are considered most ancient and can be reduced to other FeS-type configurations found in proteins (i.e., 4Fe3S, 3Fe3S, and 2Fe2S) with the removal of atoms.

To overcome issues with calculating similarity due to size and dimensionality when comparing a periodic crystal lattice to a single structure, as reported previously (Zhao et al. 2020), we reduce the crystal lattice into a periodic graph that can be directly compared to the graph representation of protein metal cofactors. For example, the periodic structure of greigite can be represented as a single structure by first, transforming the conventional unit cell (Fig. 2 left) into the Niggli reduced unit cell (Fig. 2 right), resulting in a unit cell with smaller volume (Fig. 2 center). After transformation, the reduced unit cell contains the



**Fig. 2** Transformation of the greigite conventional unit cell into the Niggli reduced unit cell. (left) The conventional unit cell (space group  $Fd\bar{3}m$  with lattice parameters  $a, b, c = 9.876$  and  $\alpha, \beta, \gamma = 90^\circ$ ) is transformed into (right) the Niggli reduced unit cell (space group  $P1$  with lattice parameters  $a, b, c = 6.983$ ,  $\alpha, \beta, \gamma = 60^\circ$ ). (center) Note the Niggli transformation (blue), for greigite, results in a reduction of volume of the conventional unit cell (green). Each unit cell is drawn with only atoms that fit entirely in the unit cells. Red, green, and blue cell edges represent the  $a$ ,  $b$ , and  $c$  axes of the unit cells

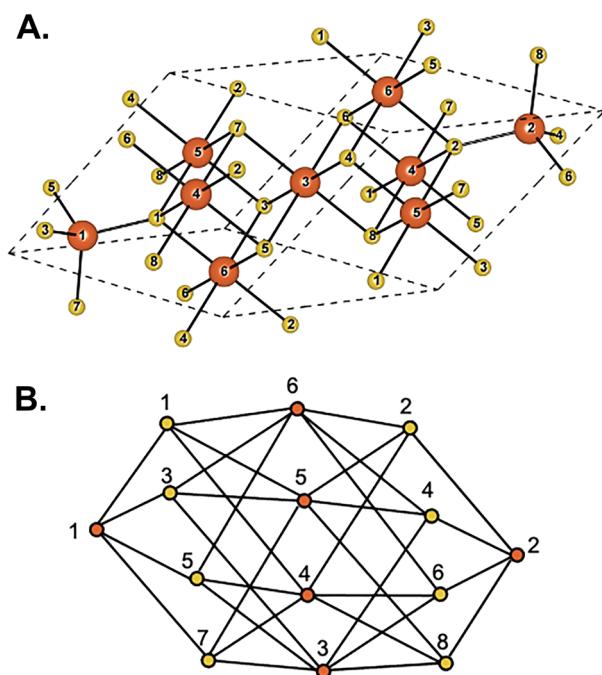
unique topological composition of atoms (Fig. 3A) needed to generate the entire crystal lattice only through translations of the reduced unit cell. The composition of the reduced unit cell represented as a quotient graph (Fig. 3B) allows the direct comparison of a periodic (infinite) to a non-periodic (finite) structure by searching for the maximum common connected edge subgraph between the mineral quotient graphs (periodic) and the protein iron sulfur cluster graphs (non-periodic).

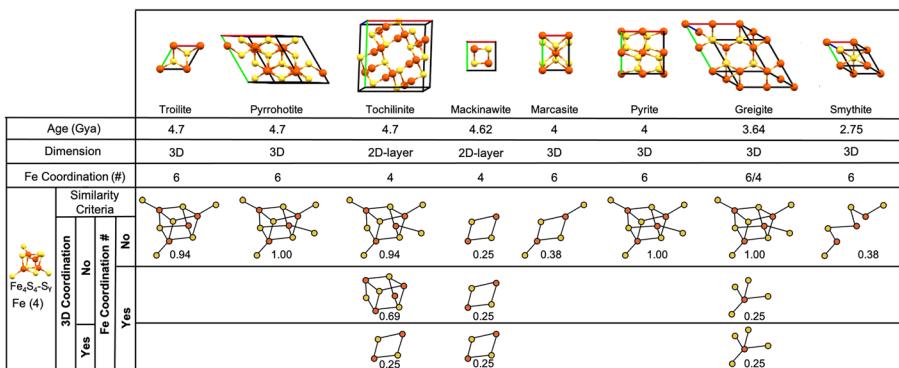
To test our comparison method, we compared the ferredoxin-type 4Fe4S cluster, including cysteine thiol sulfurs (4Fe4S-4S $\gamma$ ), to a set of iron-sulfur minerals implicated in early prebiotic chemistry and redox reactions (de Aldecoa et al. 2013; White et al. 2015) with the earliest known age of existence spanning ~ 2Gya (Fig. 4). Although the same iron-sulfur minerals are still in existence (Krivovichev et al. 2018), we expected structural similarity to encode the environmental forces within which the iron-clusters formed and therefore possibly indicate relationships between the geosphere and biosphere over time.

Comparisons between 4Fe4S-4S $\gamma$  and mineral structures were performed with increasingly stringent molecular similarity criteria from only chemical topology, to including coordination number and 3D coordination. Considering only topology or connectedness, pyrrhotite, pyrite, and greigite shared the largest connected common edge subgraph and had a Tversky index of 1 (i.e., all edges were common).

Including topology and Fe coordination number (i.e., how many atoms are directly bound), decreased molecular similarity to the ferredoxin-type Fe-S for all but mackinawite, which retained the shared topological 2Fe2S motif also in 4Fe4S-4S $\gamma$ . In addition to topology and Fe coordination number, including dimensionality of Fe coordination (i.e., 2D vs 3D lattice) led to the finding that tochilinite, mackinawite and greigite shared two common motifs with 4Fe4S-4S $\gamma$ , 2Fe2S and Fe4S, respectively (Figs. 4 and 5A, C). Finally, in addition to the previous constraints, including sulfur coordination as a cut-off variable led to no similarity or common edges between protein and mineral Fe-S clusters. Previous

**Fig. 3** The A labeled Niggli reduced unit cell of greigite and corresponding B labeled quotient graph

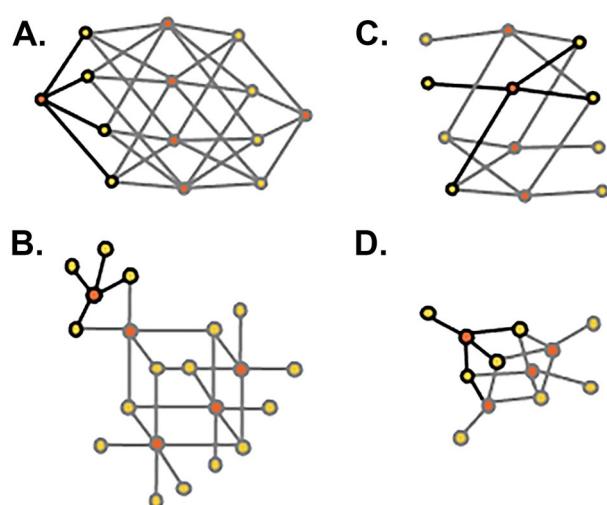




**Fig. 4** Graph similarities of mineral iron-sulfur clusters within the bulk crystal lattice and 4Fe4S-4Sy protein cluster. Niggli-reduced unit cell (top row) and a representative 4Fe4S-4Sy cluster found in proteins (bottom left column). Bottom three rows are the maximum common connected edge subgraphs (MCCES) of shared bonds (black lines) and atoms between mineral and protein cluster mapped onto the 4Fe4S-4Sy cluster given combinations of similarity criteria constraints involving 3D coordination and Fe coordination number. Values under each MCCES represent the Tversky similarity index. Orange and yellow spheres represent iron and sulfur atoms, respectively. Maximum age of minerals from <http://rruff.info/evolution>; accessed 27 July 2021

studies comparing Fe-S cofactors with proteins found in ancient active sites with the cubane 4Fe4S cluster within greigite (Russell and Hall 1997, 2006; Nitschke et al. 2013; Zhao et al. 2020), suggested extensive similarity if only topology and atom composition were considered. However, we find minimal similarity (i.e., Tversky index of 0.25). Notably, the coordination of Fe is octahedral within the cubane Fe-S of greigite and tetrahedral in the 4Fe-4S ferredoxin cofactor (Fig. 5B, D).

**Fig. 5** Maximum common connected edge subgraph of A the quotient graph of greigite and C the graph of ferredoxin 4Fe4S-4Sy as mapped onto the quotient graph of greigite, and corresponding structures of B the FeS4 which connects two FeS cubane clusters within greigite and of D the FeS4 within the FeS cubane cluster within ferredoxin, respectively. Matching edges and atoms are highlighted with black lines. Similarity criteria includes topology, coordination number, and dimensionality of mineral lattice (i.e., 2D layer vs. 3D)



## Discussion

Under the constraints of the method described in preceding sections, we demonstrate limited structural and chemical similarities between FeS species in the bulk lattice of ancient minerals and those found in proteins. This challenges the model that ancient FeS proteins directly descended from a primordial protein-mineral complex. How then did the initial protein bound iron-sulfur clusters form? Given the high likelihood of conditions on early Earth producing an iron-sulfur world (Wächtershäuser 1992) with oceans rich in soluble FeS (e.g., due to erosion (Gu et al. 2020)), iron-sulfur species likely arose spontaneously, but trapping of isolated clusters likely required peptides. Something akin to this process is recapitulated in the laboratory protocols for in vitro assembly of FeS proteins (Venkateswara Rao and Holm 2004). Associations with organic material (e.g., peptides) may have led to the observed structural differences between formed minerals and spontaneously formed or “ready-made” (Russell and Hall 2006) soluble FeS clusters. Interactions with peptides may have concurrently inhibited Fe and S reabsorption into existing mineral lattices. In vitro reconstitution of 1Fe, 2Fe2S, 4Fe4S-type clusters and binding to synthetic peptides (Kim et al. 2018; Bonfio 2021), in the absence of observed mineral formation, supports the formation of protein FeS clusters under different chemical constraints (i.e., peptide interactions) within the Hadean and Archaean oceans. In conjunction with spontaneous cluster formation and peptide interactions, formation of iron monosulfide protocells (Russell and Hall 1997; Martin and Russell 2003) may have further separated the pathways of FeS mineral formation and the chemical gradients forming biologically relevant iron-sulfur clusters.

Although limited structural similarity was observed in this study, two common substructures (2Fe2S, FeS4) are of evolutionary biogeochemical interest. In regards to the 2Fe2S substructure, mackinawite (White et al. 2015) and tochilinite (Organova et al. 1974), are 2D-layered minerals, that share the 2Fe2S motif with the ferredoxin 4Fe4S-4S<sub>y</sub> cluster. Layered minerals are thought to play a critical role in concentrating organic material in a protected environment capable of catalyzing peptide formation on early Earth (see Hazen for a review (Hazen 2006)). Periods of heavy bombardment, on early Earth by mackinawite and tochilinite containing meteorites (Buchwald 1977) may have played a role in generating organic synthesis gradients. Additionally, extant organisms use 2Fe2S motifs as a building block for higher order iron-sulfur cluster formation (Srour et al. 2020; Braymer et al. 2021) and this process may have originated in relation to the presence of mackinawite and tochilinite.

Previous observations of mineral-protein homologies are tantalizing because minerals such as greigite have been dated back to 3.64 Gya (<http://rruff.info/evolution>; accessed 27 July 2021 ID 3086), about the time interval argued for when life emerged (Pearce et al. 2018). The existence of both tetrahedral and octahedral Fe-S coordination may represent a potential shift in environmental conditions during the Archean eon that may have been influenced by life. Today, the existence of biotic greigite is due to the terminal stage of the biological Fe-S biosynthesis control pathway where cellular organisms excrete it (Gorlas et al. 2018).

Low similarities between FeS species in the bulk mineral lattice to protein FeS clusters does not necessarily obviate their evolutionary connection. Our analysis highlights potentially more revealing approaches that may uncover evolutionary connections between FeS species in the bulk mineral lattice to protein FeS clusters. Although greigite has been considered evolutionarily similar to ferredoxin-type proteins because it is able to catalyze

analogous pre-biotic reactions (e.g., CO<sub>2</sub> reduction (Roldan et al. 2015)), it is actually the surface of greigite catalyzing these chemical reactions, not the bulk solid. Rather than comparing protein iron-sulfur clusters and the bulk lattice structure of iron-sulfide minerals, similarities to relaxation bonding environments on mineral surfaces should be investigated (Zieliński 1996). The real challenge will be comparing mineral surfaces to protein active sites (Vaughan et al. 1997). However, this is hampered by the complexity of mineral surfaces: complex 3D bumps, pits, and edges, each with a different cation topology, especially along the edges, which are known to have different coordination numbers, distances, and angles from the bulk solid (Waychunas and Zhang 2008). Tying the relationship between mineral surface chemistry and the step-by-step processes extant organisms use to extract metals from the environment and assemble functional metal-clusters (Srour et al. 2020; Braymer et al. 2021) will likely be a productive way to develop a framework for protein-mineral co-evolution.

Lastly, Fe-S cluster formation is inherently driven by the presence of minerals, however absolute differences in structure between mineral and protein iron-sulfur clusters may be driven by the metabolic needs of the cell. For instance, peptide-FeS interactions enabling hydrogen bonding (Lin et al. 2005) may have allowed the range of chemical potentials required for life to persist and maintain disequilibrium with the environment. In other words, peptides likely molded or modified the mineral elements to meet the thermodynamic demands of life.

## Conclusion

Graph theoretic approaches allowed us to compare the infinite mineral lattice structurally and chemically to the finite protein metal cluster. Maximum common connected edge subgraphs not including coordination number of the bulk FeS clusters within the mineral lattice and the individual FeS clusters within proteins show great similarity between metal clusters. However, including coordination number significantly reduces substructure similarity. Shared common substructures (2Fe2S and Fe4S motifs) may have arisen from ancient environmental constraints on geological and biological processes. This approach can easily be extended to the complete set of metalloproteins and minerals, extending this analysis beyond iron and sulfur species. We note that while mineral structures provide significant insight into geochemistry through time, a more relevant aspect of study may be mineral surfaces, where functional connections with proteins are more appropriate. Informatic approaches such as those proposed here, combined with computational chemical modeling may allow us to develop a comprehensive framework describing the chemical evolution of proteins and minerals over deep time.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11084-022-09630-x>.

**Acknowledgements** The authors wish to acknowledge Robert Hazen, Saroj Poudel, Corday Selden, Jennifer Timm, Alexei Tyryshkin and Jan Seiss for useful discussions. The authors would also like to thank the reviewers for critically reading this manuscript and providing exceptional feedback that helped produce a more thorough manuscript. In addition, the authors would like to send their greatest appreciation for the scientists that meticulously collected and deposited mineral and protein structural data into public repositories.

**Funding** This work was supported by NASA grant 80NSSC18M0093 from the NASA Astrobiology Institute. K.N.M. acknowledges support from the NIH IRACDA Postdoctoral Fellowship Program Grant K12GM093854.

**Data and Software Availability** Source code of comparing two molecular graphs based on their maximum common connected edge subgraph, examples and supporting material are freely available at: <https://github.com/AlBi-HHU/como>.

## Declarations

**Conflict of Interest** No potential conflict of interest was reported by the authors.

## References

- Andreini C, Banci L, Bertini I et al (2007) Non-heme iron through the three domains of life. *Proteins Struct Funct Bioinforma* 67:317–324. <https://doi.org/10.1002/prot.21324>
- Andreini C, Bertini I, Cavallaro G et al (2009) Structural analysis of metal sites in proteins: Non-heme iron sites as a case study. *J Mol Biol* 388:356–380. <https://doi.org/10.1016/j.jmb.2009.02.052>
- Andreini C, Cavallaro G, Lorenzini S (2012) FindGeo: A tool for determining metal coordination geometry. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bts246>
- Andrews LC, Bernstein HJ, Sauter NK (2019) Selling reduction versus Niggli reduction for crystallographic lattices. *Acta Crystallogr Sect A Found Adv* 75:115–120. <https://doi.org/10.1107/S2053273318015413>
- Bahiense L, Manić G, Piva B, De Souza CC (2012) The maximum common edge subgraph problem: A polyhedral investigation. *Discret Appl Math* 160:2523–2541. <https://doi.org/10.1016/j.dam.2012.01.026>
- Berman HM (2000) The protein data bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
- Blatov VA, Shevchenko AP, Serezhkin VN (2000) TOPOS3.2: A new version of the program package for multipurpose crystal-chemical analysis. *J Appl Crystallogr* 33:1193. <https://doi.org/10.1107/S0021889800007202>
- Bonfio C (2021) The curious case of peptide-coordinated iron-sulfur clusters: Prebiotic and biomimetic insights. *Dalt Trans* 50:801–807. <https://doi.org/10.1039/d0dt03947k>
- Braymer JJ, Freibert SA, Rakwalska-Bange M, Lill R (2021) Mechanistic concepts of iron-sulfur protein biogenesis in Biology. *Biochim Biophys Acta - Mol Cell Res* 1868:118863. <https://doi.org/10.1016/j.bbamcr.2020.118863>
- Buchwald V (1977) The mineralogy of iron meteorites. *Philos Trans R Soc London Ser Math Phys Sci* 286:453–491. <https://doi.org/10.1098/rsta.1977.0127>
- Chung SJ, Hahn T, Klee WE (1984) Nomenclature and generation of three-periodic nets: the vector method. *Acta Crystallogr Sect A Found Crystallogr* 40:42–50. <https://doi.org/10.1107/S0108767384000088>
- Cock PJA, Antao T, Chang JT et al (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- de Aldecoa ALI, Roldán FV, Menor-Salván C (2013) Natural pyrrhotite as a catalyst in prebiotic chemical evolution. *Life* 3:502–517. <https://doi.org/10.3390/life3030502>
- Downs RT, Hall-Wallace M (2003) The American Mineralogist crystal structure database. *The American mineralogist* 88:247–250
- Gelato LM, Parthé E (1987) STRUCTURE TIDY— a computer program to standardize crystal structure data. *J Appl Crystallogr* 20:139–143. <https://doi.org/10.1107/S0021889887086965>
- Gorlas A, Jacquemot P, Guigner JM et al (2018) Greigite nanocrystals produced by hyperthermophilic archaea of Thermococcales order. *PLoS ONE* 13:1–10. <https://doi.org/10.1371/journal.pone.0201549>
- Gu X, Heaney PJ, Reis FDAA, Brantley SL (2020) Deep abiotic weathering of pyrite. *Science* 23:eabb8092. <https://doi.org/10.1126/science.abb8092>
- Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using networkX. Proceedings of the 7th Python in Science Conference, pp 11–15
- Hazen RM (2006) Mineral surfaces and the prebiotic selection and organization of biomolecules. *Am Mineral* 91:1715–1729. <https://doi.org/10.2138/am.2006.2289>
- Kim JD, Pike DH, Tyryshkin AM et al (2018) Minimal Heterochiral de Novo Designed 4Fe-4S Binding Peptide Capable of Robust Electron Transfer. *J Am Chem Soc* 140:11210–11213. <https://doi.org/10.1021/jacs.8b07553>

- Klee WE (2004) Crystallographic nets and their quotient graphs. *Cryst Res Technol* 39:959–968. <https://doi.org/10.1002/crat.200410281>
- Krivovichev S (2012a) Topological complexity of crystal structures: Quantitative approach. *Acta Crystallogr Sect A Found Crystallogr* 68:393–398. <https://doi.org/10.1107/S0108767312012044>
- Krivovichev SV (2013) Structural complexity of minerals: information storage and processing in the mineral world. *Mineral Mag* 77:275–326. <https://doi.org/10.1180/minmag.2013.077.3.05>
- Krivovichev SV (2012b) Information-based measures of structural complexity: Application to fluorite-related structures. *Struct Chem* 23:1045–1052. <https://doi.org/10.1007/s11224-012-0015-1>
- Krivovichev SV, Krivovichev VG, Hazen RM (2018) Structural and chemical complexity of minerals: correlations and time evolution. *Eur J Mineral* 30:231–236. <https://doi.org/10.1127/ejm/2018/0030-2694>
- Křivý I, Gruber B (1976) A unified algorithm for determining the reduced (Niggli) cell. *Acta Crystallogr Sect A* 32:297–298. <https://doi.org/10.1107/S0567739476000636>
- Kuppuraj G, Dudev M, Lin C (2009) Factors governing metal-ligand distances and coordination geometries of metal complexes. *J Phys Chem B* 113:2952–2960. <https://doi.org/10.1021/jp807972e>
- Lin IJ, Gebel EB, Machonkin TE et al (2005) Changes in hydrogen-bond strengths explain reduction potentials in 10 rubredoxin variants. *Proc Natl Acad Sci U S A* 102:14581–14586. <https://doi.org/10.1073/pnas.0505521102>
- Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc London Ser B Biol Sci* 358:59–85. <https://doi.org/10.1098/rstb.2002.1183>
- Mathé C, Weill CO, Mattioli TA et al (2007) Assessing the role of the active-site cysteine ligand in the superoxide reductase from desulfovibrio baarsii. *J Biol Chem* 282:22207–22216. <https://doi.org/10.1074/jbc.M700279200>
- Momma K, Izumi F (2011) VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J Appl Crystallogr*. <https://doi.org/10.1107/S0021889811038970>
- Moriarty NW, Adams PD (2019) Iron–sulfur clusters have no right angles. *Acta Crystallogr Sect D Struct Biol* 75:16–20. <https://doi.org/10.1107/S205979831801519X>
- Müller P, Köpke S, Sheldrick GM (2003) Is the bond-valence method able to identify metal atoms in protein structures? *Acta Crystallogr Sect D Biol Crystallogr* 59:32–37. <https://doi.org/10.1107/S0907444902018000>
- Nitschke W, McGlynn SE, Milner-White EJ, Russell MJ (2013) On the antiquity of metalloenzymes and their substrates in bioenergetics. *Biochim Biophys Acta - Bioenerg* 1827:871–881. <https://doi.org/10.1016/j.bbabiobioenerg.2013.02.008>
- Novikov Y, Copley SD (2013) Reactivity landscape of pyruvate under simulated hydrothermal vent conditions. *Proc Natl Acad Sci U S A* 110:13283–13288. <https://doi.org/10.1073/pnas.1304923110>
- Organova NI, Drits VA, Dmitrik AL (1973) Structural study of tochilinite-II. Acicular variety unusual diffraction patterns. *Kristallografiya* 18:966–972
- Organova NI, Drits VA, Dmitrik AL (1974) Structural study of tochilinite. *Sov Phys Crystallogr* 18:606–609
- Gurobi Optimization LLC (2021) Gurobi optimizer reference manual
- Pearce BKD, Tupper AS, Pudritz RE, Higgs PG (2018) Constraining the time interval for the origin of life on earth. *Astrobiology* 18:343–364. <https://doi.org/10.1089/ast.2017.1674>
- Putignano V, Rosato A, Banci L, Andreini C (2018) MetalPDB in 2018: A database of metal sites in biological macromolecular structures. *Nucleic Acids Res* 46:D459–D464. <https://doi.org/10.1093/nar/gkx989>
- Raanan H, Pike DH, Moore EK et al (2018) Modular origins of biological electron transfer chains. *Proc Natl Acad Sci* 115:201714225. <https://doi.org/10.1073/pnas.1714225115>
- Raanan H, Poudel S, Pike DH et al (2020) Small protein folds at the root of an ancient metabolic network. *Proc Natl Acad Sci* 117:7193–7199. <https://doi.org/10.1073/pnas.1914982117>
- Rickard D, Luther GW (2007) Chemistry of iron sulfides. *Chem Rev* 107:514–562. <https://doi.org/10.1021/cr0503658>
- Roldan A, Hollingsworth N, Roffey A et al (2015) Bio-inspired CO<sub>2</sub> conversion by iron sulfide catalysts under sustainable conditions. *Chem Commun* 51:7501–7504. <https://doi.org/10.1039/c5cc02078f>
- Russell MJ, Hall AJ (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J Geol Soc London* 154:377–402. <https://doi.org/10.1144/gsjgs.154.3.0377>
- Russell MJ, Hall AJ (2006) The onset and early evolution of life. In: Evolution of Early Earth's Atmosphere, Hydrosphere, and Biosphere - Constraints from Ore Deposits. Geological Society of America, pp 1–32. [https://doi.org/10.1130/2006.1198\(01\)](https://doi.org/10.1130/2006.1198(01))

- Santoro A, Michell AD (1970) Determination of reduced cells. *Acta Crystallogr Sect A* 26:124–127. <https://doi.org/10.1107/S0567739470000177>
- Srou B, Gervason S, Monfort B, D'autréaux B (2020) Mechanism of iron–sulfur cluster assembly: In the intimacy of iron and sulfur encounter. *Inorganics* 8:1–38. <https://doi.org/10.3390/inorganics8100055>
- Tversky A (1977) Features of similarity. *Psychol Rev* 84:327–352. <https://doi.org/10.1037/0033-295X.84.4.327>
- Valasatava Y, Rosato A, Furnham N et al (2018) To what extent do structural changes in catalytic metal sites affect enzyme function? *J Inorg Biochem* 179:40–53. <https://doi.org/10.1016/j.jinorgbio.2017.11.002>
- Vaughan DJ, Becker U, Wright K (1997) Sulphide mineral surfaces: Theory and experiment. *Int J Miner Process* 51:1–14. [https://doi.org/10.1016/s0301-7516\(97\)00035-5](https://doi.org/10.1016/s0301-7516(97)00035-5)
- Venkateswara Rao P, Holm RH (2004) Synthetic Analogues of the Active Sites of Iron-Sulfur Proteins. *Chem Rev* 104:527–559. <https://doi.org/10.1021/cr020615+>
- Wächtershäuser G (1992) Groundworks for an evolutionary biochemistry: The iron-sulphur world. *Prog Biophys Mol Biol* 58:85–201. [https://doi.org/10.1016/0079-6107\(92\)90022-X](https://doi.org/10.1016/0079-6107(92)90022-X)
- Waychunas GA, Zhang H (2008) Structure, chemistry, and properties of mineral nanoparticles. *Elements* 4:381–387. <https://doi.org/10.2113/gselements.4.6.381>
- White LM, Bhartia R, Stucky GD et al (2015) Mackinawite and greigite in ancient alkaline hydrothermal chimneys: Identifying potential key catalysts for emergent life. *Earth Planet Sci Lett* 430:105–114. <https://doi.org/10.1016/j.epsl.2015.08.013>
- Williams RJP (1981) The Bakerian Lecture, 1981 Natural Selection of the Chemical elements. *Proc R Soc London Ser B Biol Sci* 213:361–397. <https://doi.org/10.1098/rspb.1981.0071>
- Zagorac D, Muller H, Ruehl S et al (2019) Recent developments in the Inorganic Crystal Structure Database: Theoretical crystal structure data and related features. *J Appl Crystallogr* 52:918–925. <https://doi.org/10.1107/S160057671900997X>
- Zhao D, Bartlett S, Yung YL (2020) Quantifying Mineral-Ligand Structural Similarities: Bridging the Geological World of Minerals with the Biological World of Enzymes. *Life* 10:338. <https://doi.org/10.3390/life10120338>
- Zheng H, Langner KM, Shields GP et al (2017) Data mining of iron(II) and iron(III) bond-valence parameters, and their relevance for macromolecular crystallography. *Acta Crystallogr Sect D Struct Biol* 73:316–325. <https://doi.org/10.1107/S2059798317000584>
- Zielinski P (1996) Review of surface relaxation and reconstruction phenomena. *Acta Phys Pol A* 89:251–263. <https://doi.org/10.12693/APhysPolA.89.251>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.